ANALYSIS OF PIANO TONES USING AN INHARMONIC INVERSE COMB FILTER

*Heidi-Maria Lehtonen**

Department of Signal Processing and Acoustics,
Helsinki University of Technology, TKK
Espoo, Finland
heidi-maria.lehtonen@tkk.fi

ABSTRACT

This paper presents a filter configuration for canceling and separating partials from inharmonic piano tones. The proposed configuration is based on inverse comb filtering, in which the delay line is replaced with a high-order filter that has a proper phase response. Two filter design techniques are tested with the method: an FIR filter, which is designed using frequency sampling, and an IIR filter, which consists of a set of second-order allpass filters that match the desired group delay. It is concluded that it is possible to obtain more accurate results with the FIR filter, while the IIR filter is computationally more efficient. The paper shows that the proposed analysis method provides an effective and easy way of extracting the residual signal and selecting partials from piano tones. This method is suitable for analysis of recorded piano tones.

1. INTRODUCTION

In this paper, the extraction of the residual signal and selection of single partials from inharmonic piano tones by means of digital filtering are discussed. Both the tonal and noise components are essential in signal analysis. For example, the decay rates of the partials [1] or analysis of the sustain pedal [2, 3] provide important information for sound synthesis. The proposed residual signal extraction technique is based on inverse comb filtering, an approach that was used already by Moorer in the 1970s for pitch detection in speech signals [4], and analysis of musical tones for obtaining parameters for additive synthesis [5]. In this work, the delay line in the inverse comb filter is replaced by an FIR or IIR filter, which has a proper phase specification for matching the partial frequencies of an inharmonic piano tone. Recently, this idea has been used for analysis of harmonic musical instrument tones [6, 7, 8], where the spectral components are integral multiples of the fundamental frequency. This is not the case with the piano, however, where the partial components are stretched upwards in frequency due to dispersion, a typical feature of stiff musical instrument strings.

The purpose of this work is to provide a simple yet effective tool for analyzing inharmonic musical instrument tones using the same starting point as in [6, 7, 8]. The required parameters for the method are the fundamental frequency and the inharmonicity coefficient. These parameters can be obtained by using some inharmonicity estimation algorithm, such as the inharmonic comb filter method proposed by Galembo and Askenfelt [9] or the partial frequency deviation algorithm presented by Rauhala *et al.* [10]. Recently, Wang and Tan presented a method based on wavelets for analyzing the dispersion [11].

The dispersion phenomenon is clearly audible in low piano tones, and it must be taken into account in high-quality sound synthesis. Different methods for modeling the dispersive behavior of the piano strings have been proposed in the literature [12, 13, 14, 15, 16, 17]. Most of the previous dispersion filter design techniques are used for real-time synthesis purposes, and they are able to roughly approximate the dispersion phenomenon; this is sufficient, since the human hearing system is not extremely sensitive to accuracy in inharmonicity [18, 19, 20]. In analysis of recorded piano tones, more precision is required, however. In this paper, a large FIR filter and an IIR filter proposed by Abel and Smith [17] are used to accurately model the dispersive behavior in the inverse comb filter.

Other methods for extracting the residual signal have been proposed earlier. Sinusoidal modeling [21, 22, 23] is a method that is based on the analysis of the target signal in frames with windowed FFT, where spectral peaks are picked in order to obtain frequency, amplitude, and phase information for each partial. The data obtained from each frame are connected in adjacent data points to form frequency, amplitude, and tracks, which are used in synthesizing the desired signal component. Sinusoidal modeling is a powerful analysis method, especially for the analysis of very complex time-varying signals. Methods that apply inverse filtering in some form include the matrix pencil inverse filtering [24] and inverse filtering with sinusoids plus noise [25]. Recently, Lee *et al.* proposed a statistical spectral interpolation method for extracting the excitation signal from plucked guitar tones [26]. Other possibilities for analyzing the spectral contents of musical instruments include wavelets [27, 11], high-resolution tracking methods [28, 29], and frequency-zooming ARMA modeling [30, 31]. An advantage of the filtering-based analysis tool over other, more sophisticated analysis methods is its simplicity; the filter can be designed based on the fundamental frequency and the inharmonicity coefficient value, and after the design process the analysis can be carried out with a single filtering operation.

The residual signal has an essential part in physics-based sound synthesis. In digital waveguide modeling [32, 33], the residual signal can be obtained by filtering the desired output with the inverted transfer function of the desired string model [1]. This signal can then be used as an excitation signal for the synthesis algorithm. In order to conduct inverse filtering for obtaining the excitation signal, the partial decay times should be taken into account in the filter design in addition to the dispersion phenomenon. Inverse filtering for this purpose is not considered in the present work, however, the main reason being the complicated decay process of the piano tones. Dealing with the decay-related features that are character-

* This work was supported by the GETA Graduate School, the Academy of Finland (project no. 122815), the Finnish Cultural Foundation, the Nokia Foundation, and Emil Aaltonen Foundation.

istics of the piano tones, such as the two-stage decay and beating [34] accurately enough would require a separate filter for matching the loop gain values. Moreover, since the goal in the present work is to analyze the partials and residuals of piano tones, matching the loop gain values is not of primary interest.

This paper is organized as follows. First, in Sec. 2 the structure of the inverse comb filter is presented, and the replacement of the delay line with two alternative dispersion filters, an FIR filter designed with frequency sampling and an IIR filter proposed by Abel and Smith [17], is discussed. After this, the performance of the two dispersion filters as a part of the inverse comb filter are compared in Sec. 3. Section 4 illustrates how the proposed method works in practice: single partial components are selected from synthetic tones, and residual signals are extracted from two recorded piano tones. Finally, conclusions are drawn in Sec. 5. Sound examples are provided at http://www.acoustics.hut.fi/go/DAFx08-IICF/.

2. INHARMONIC INVERSE COMB FILTERS

The inverse comb filter (ICF) is an FIR filter in which the delayed version of the input signal is subtracted from the original input signal. The block diagram of an ICF is shown in Fig. 1 (a). The transfer function of such a system can be written as

$$H_{\mathrm{ICF}}(z) = \frac{1}{2}(1 - z^{-L}), \tag{1}$$

where L is the length of the delay line and the coefficient $1/2$ sets the gain to unity in the passband. The magnitude response of the ICF contains notches at nf_{s}/L, where $n \in \mathbb{Z}$ and f_{s} is the sampling rate in Hz. Since the delay line length L in Eq. (1) is restricted to be an integer, it might happen that the filter is not capable of effectively canceling the harmonic components. For example, if the fundamental frequency of the tone is $f_0 = 261.626$ Hz (which corresponds to the middle C in the piano), the required delay line length would be $L = f_{\mathrm{s}}/f_0 = 168.562$ samples, when the sampling rate of 44.1 kHz is used. In practice, the delay line in Fig. 1 (a) can be replaced with a fractional-delay filter [35, 36, 37], which fine-tunes the delay line length so that the notches of the filter appear exactly at the harmonic frequencies. This is illustrated in Fig. 1 (b), in which $H(z)$ is the fractional-delay filter. Earlier studies [6, 7, 8] show that fractional-delay ICFs can be used to cancel partial components effectively, when the musical tone is strictly harmonic. In the case of the piano, the harmonic frequencies are not exactly integral multiples of the fundamental frequency. Thus, canceling partials from a piano tone requires a filter, which has a nonlinear phase specification.

2.1. Design of Dispersive FIR Filter

The phase response of the dispersion filter is defined from the knowledge that the phase is a multiple of 2π radians at every partial frequency. In the case of the piano, the partial frequencies can be obtained from Eq. (2) [34] that follows:

$$f_k = kf_0\sqrt{1 + k^2 B}, \tag{2}$$

where k refers to the partial index, f_0 is the nominal fundamental frequency of the tone, and B is the inharmonicity coefficient. Thus, the value of the phase response $\theta(\omega)$ is $-2\pi k$ at frequency f_k. In the case of a strictly harmonic tone, the coefficient B is equal to 0 and the partial frequencies are integral multiples of the nominal fundamental frequency, which leads to a linear-phase filter. When $B > 0$, the partials are displaced upwards in frequency, and the phase response becomes nonlinear. Since the phase response $\theta(\omega)$ is assumed to be a smooth curve, cubic spline interpolation is performed between the partial frequencies so that the frequency response can be computed as a continuous function of frequency. The frequency response of the dispersive FIR filter can now be written as

$$H(\omega) = |H(\omega)|e^{j\theta(\omega)}. \tag{3}$$

Since it is impossible to design a dispersive allpass FIR filter, it is proposed that the magnitude response imitates a lowpass filter, whose passband covers the range $0 - 20$ kHz. The remaining part is defined by means of a raised cosine function (see Fig. 2(a)). The cosine function is a natural choice for the transition and stopbands, since this specification is easily realizable with FIR filters [38]. The approximation bandwidth $0 - 20$ kHz is suitable for analyzing most musical instrument tones at the sampling rate of 44.1 kHz, because in recorded tones the remaining frequency band above 20 kHz contains little or no information of interest. Moreover, the upper limit of the range of human hearing is about 20 kHz.

The magnitude and phase delay specifications, and the impulse response of the filter are shown in Figs. 2(a), 2(b), and 3, respectively, when the nominal fundamental frequency f_0 and the inharmonicity coefficient B are set to be 261.6 Hz and 3.0×10^{-4}, respectively.

The impulse response, and thus the FIR filter coefficients, can be obtained from the frequency response of Eq. (3) by applying the inverse discrete Fourier transform. When the number of data

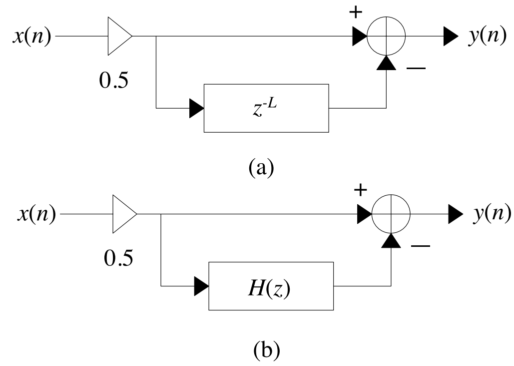

Figure 1: (a) Block diagram of the conventional ICF and (b) an ICF with a fractional-delay filter $H(z)$ (after [6]).

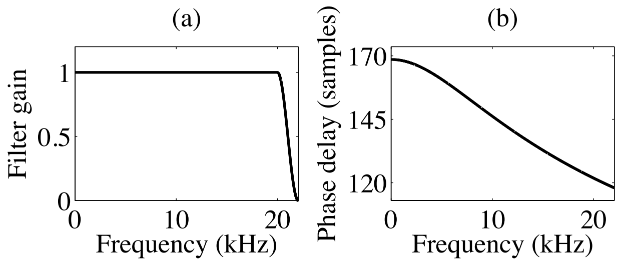

Figure 2: (a) The magnitude and (b) phase delay specifications of the dispersive FIR filter ($f_0 = 261.6$ Hz, $B = 3.0 \times 10^{-4}$). In (c), the truncated impulse response of length 200 samples is shown.

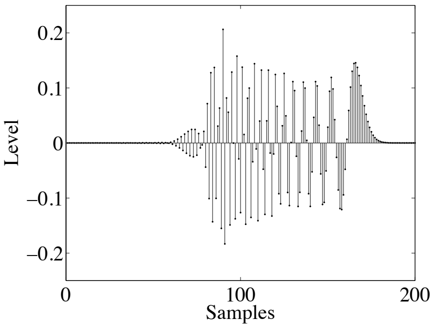

Figure 3: Truncated impulse response of the dispersive FIR filter of length 200 samples ($f_0 = 261.6$ Hz, $B = 3.0 \times 10^{-4}$).

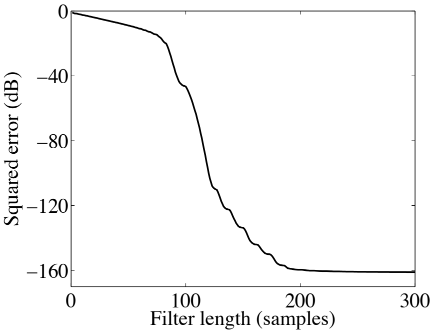

Figure 4: Squared error of the truncated impulse response as a function of filter length when $f_0 = 261.6$ Hz and $B = 3.0 \times 10^{-4}$.

points in the frequency response vector is large, say, 44100 points, good accuracy is obtained. The length of the impulse response can be decreased by truncating the prototype impulse response with a rectangular window function. By choosing the window position so that the squared sum of the truncated impulse response is maximized, the performance of the truncated filter can be optimized in terms of the least-squares error. It is also possible to use some other window function, such as the Hamming window. In practice, however, in order to achieve good accuracy, the response should be truncated so that the largest samples are included in the final response, indicating that the excluded sample values are small in any case. In addition, the rectangular window is known to be optimal in the least-squares sense [38].

The position that corresponds to the maximum squared sum can be found by sliding the rectangular window of a certain length over the prototype impulse response. When the best position is found, the truncated impulse response can be cascaded, if needed, with a delay line in order to set the phase delay at the fundamental frequency to the correct value. In Fig. 4 the squared error of the impulse response is displayed as a function of filter length. As can be seen, the approximation error decreases monotonously with filter length and saturates at some point, depending on the f_0 and B parameters corresponding to the filter specification. In this case, the saturation point is found at approximately 190 samples (see Fig. 4).

When the delay line of an ICF is replaced by the proposed dispersive FIR filter (see Fig. 1 (b)), the notches of the ICF occur

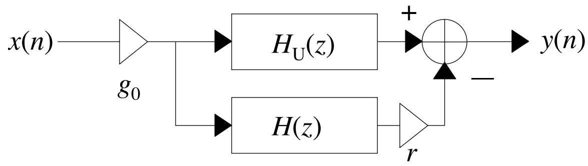

Figure 5: Block diagram of the FIR inharmonic inverse comb filter.

at the partial frequencies of an inharmonic tone. In order to obtain an accurate design, the phase response of the whole structure is approximated instead of the phase response of the dispersion filter. The corresponding block diagram is shown in Fig. 5. The filter H_{U} in the upper branch is used to synchronize the signals, and it has the same magnitude response as the dispersion filter (3) and a linear phase response. Thus, the frequency response of the structure can be written as

$$H_{\mathrm{IICF}}(\omega) = g_0[H_{\mathrm{U}}(\omega) - rH(\omega)], \tag{4}$$

where the coefficients g_0 and r are used to set the maximum gain to unity and the gain at the bottom of the notches to a desired level, such as 10^{-6}, which corresponds to 120 dB attenuation at frequencies f_k. When the attenuation in dB is denoted with A_{dB}, the g_0 and r parameters can be computed as [7]

$$A = 10^{\frac{-A_{\mathrm{dB}}}{20}}, \tag{5}$$

$$r^L = \frac{1-A}{1+A}, \tag{6}$$

$$r = \sqrt[L]{\frac{1-A}{1+A}}, \tag{7}$$

$$g_0 = \frac{1}{1+r^L}. \tag{8}$$

Since the analysis is conducted off-line and it is important to achieve accurate results, the computational cost is not a critical issue. For example, for the tone C1 ($f_0 = 32.7$ Hz, $B = 2.62 \times 10^{-4}$) a filter length 2050 is found to be adequate for achieving a 100 dB attenuation at partial frequencies.

2.2. Dispersive IIR Filter Design

Recently, Abel and Smith proposed a nonparametric allpass filter design technique, which is capable of matching the desired group delay specification accurately [17]. The method is also numerically robust, even with high filter orders, unlike many other allpass filter design techniques that try to match a certain group delay specification. The idea is to divide the desired group delay characteristics into sections, which when integrated along the frequency axis have an area of 2π, and to assign a pole-zero pair to each section. These pole-zero pairs, when arranged in allpass form, will contribute exactly 2π to the total group delay of the filter. The filter order is determined from the number of 2π integrated delay sections. The filter order can be decreased, especially in the case of low fundamental frequencies, by implementing a part of the desired delay with a delay line.

After dividing the group delay into sections with integrated area of 2π, a first-order complex allpass filter

$$G(z) = \frac{-\rho e^{-j\theta} + z^{-1}}{1 - \rho e^{j\theta} z^{-1}} \quad (9)$$

is assigned for each section. In Eq. (9) the parameters ρ and θ are the pole radius and frequency, respectively. The pole frequency is set equal to the midpoint of the corresponding frequency band $[\omega_-, \omega_+]$, and the pole radius depends on the parameter β so that the group delay at the frequency band edges is a fraction β of the group delay peak. The definition of parameters is discussed in [17]. In order to obtain filters with real coefficients, the complex allpass sections can be combined to biquads that have real coefficients.

2.2.1. *Aspects on Accuracy and Computational Cost*

The value of the parameter β is a very critical issue when the goal is to model the phase characteristics as accurately as possible. Abel and Smith point out that by selecting β values close to one it is possible to obtain smooth delay curves, while smaller values produce ripple in the delay [17]. On the other hand, with larger β values the tracking error is larger near the Nyquist limit. In the case of signal analysis, it is more important to obtain as smooth and accurate delay curves as possible, especially in the important frequency band, say $0 - 20$ kHz, while the accuracy near the Nyquist limit is not a critical issue. The author has found that $\beta = 0.95$ produces the best results in most of the design tasks. However, there is still a minor overall tracking error, also in the low frequencies, and this tracking error is proposed to be compensated with a fractional-delay tuning filter. The tuning filter can be designed, e.g. with the Thiran allpass filter design method [39, 35], which produces a maximally flat group delay. The fractional-delay part can be determined from the mean error between the target group delay specification and the group delay of the obtained filter in the frequency range $0 - 18$ kHz. After 18 kHz the tracking error is so large that it should not be taken into account.

In [17], Abel and Smith suggest that a constant delay can be added to the frequency-dependent delay specification so that it integrates to an integer multiple of 2π. This procedure adds, however, error to the group delay that is used as a specification in the design process. It is suggested here that the extra delay can be added to the target delay using a linear ramp so that the error is small at low frequencies and larger at high frequencies. This procedure ensures that the important frequency range is modeled as accurately as possible.

In order to reduce the number of second-order sections, part of the desired delay is modeled with a delay line. The length of the delay line can be determined from the target group delay at the Nyquist limit. For example, if the target group delay at the Nyquist limit is 120 samples (see e.g. Fig. 2 (b)), a delay line of length 100 samples can be used. In practice, the use of a delay line length that is equal to the target group delay at the Nyquist limit in not recommended, since the tracking error occurring with large β values may lead to negative group delay values at high frequencies. The optimization of the length of the delay line could be performed similarly as suggested by Rocchesso and Scalcon [13].

2.3. Selection of Single Harmonics

In addition to canceling all partials from a piano tone, single harmonics can be extracted as separate signals. This is implemented by cascading a second-order all-pole filter with an inharmonic inverse comb filter [6, 7, 8]. This kind of filter is called the harmonic extraction filter (HEF) [7]. In practice, an HEF filter is obtained when the pole of the second-order all-pole filter located at the frequency of the chosen partial cancels the corresponding zero of the inverse comb filter. In order to guarantee stability, it is not reasonable to set a pole on the unit circle in the z plane, but to move the pole slightly inside the unit circle. Consequently, the zeros in the transfer function of the inverse comb filter are moved inside the unit circle as well. More extensive discussion of the HEF filter design along with the choice of parameter values is given in [7].

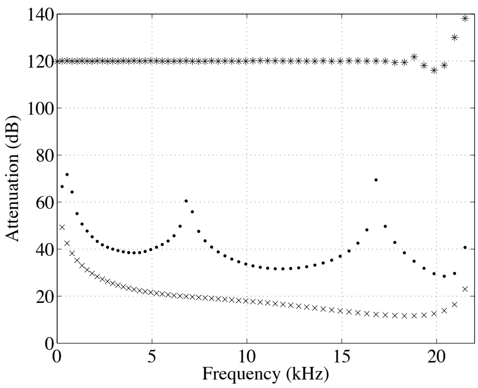

Figure 6: Comparison between the FIR dispersion filter ('*'), IIR dispersion filter with a tuning filter of order 30 ('.'), and IIR dispersion filter without the tuning filter ('x').

3. COMPARISON OF THE TWO DISPERSION FILTERS

Figure 6 compares the obtained attenuation at partial frequencies in the case of the three ICFs with an FIR and IIR dispersion filters, respectively. The asterisks ('*') represent the result in the case of the FIR filter. The dots ('.') and crosses ('x') represent the performance of the IIR filter when a tuning filter is used and when it is not used, respectively. In all cases, the parameters f_0 and B are set to $f_0 = 261.6256$ Hz and $B = 2.9936 \times 10^{-4}$. The order of the FIR dispersion filter is 1000, and the required attenuation at the partial frequencies is 120 dB. The IIR dispersion filter consists of 30 biquads and an allpass tuning filter of order 29. The β parameter is equal to 0.95 and the length of the delay, which is modeled with pure delay, is 60 samples.

It can be seen that the FIR filter clearly outperforms the IIR filter, since much better attenuation is obtained. On the other hand, the computational complexity of the FIR filter is considerably larger. However, off-line FIR filtering can be implemented efficiently using FFT convolution. Increasing the order of the IIR filter does not have an effect on accuracy, since the method automatically determines the order of the filter when the group delay specification is divided into sections of 2π. The tuning filter improves the performance of the IIR filter by $20 - 60$ dB. The peaks that are present in Fig. 6 come from the fact that the designed group delay ripples around the target group delay specification. If the fractional delay filter is excluded, the approximated delay is larger than the target delay throughout the frequency band.

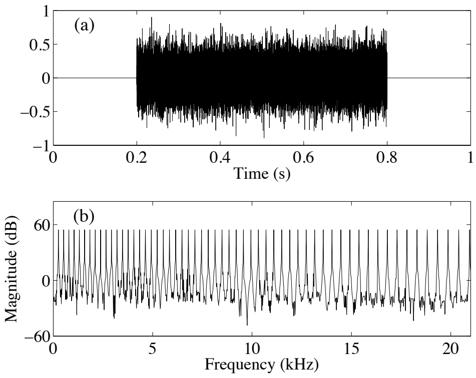

Figure 7: Inharmonic synthetic test signal (a) in the time and (b) the frequency domain.

4. APPLICATION EXAMPLES

This section shows how the inharmonic inverse comb filter copes with synthetic test tones and recorded piano tones. Two application cases are demonstrated: single harmonic components are selected by canceling the rest of the partials and the residual signal is extracted from tones. The two suggested filters, the FIR filter, and the IIR filter by Abel and Smith [17] are used as a part of the inverse comb filter. The sampling frequency is 44.1 kHz in all test cases.

4.1. Synthetic Test Tones

The inharmonic inverse comb filter is first tested with synthetic tones. The signal that is to be analyzed is the sum of sinusoids:

$$x(n) = \sum_{k=1}^{K} \sin(\frac{2\pi n f_k}{f_s} + \phi_k), \tag{10}$$

where K is the total number of harmonics included in the signal and f_k and ϕ_k are the frequency and the phase of the kth partial, respectively. The partial frequencies can be computed with Eq. (2). In this case, the fundamental frequency and the inharmonicity coefficient were set to $f_0 = 261.6256$ Hz and $B = 2.9936 \times 10^{-4}$, respectively. Using these parameters there will be $K = 58$ partial components in the frequency range $0 - 22.05$ kHz. The initial phases ϕ_k are uniformly distributed random numbers in the range $[0, 2\pi]$. The test signal is presented in Fig. 7 in the time and frequency domains. The spectrum is computed from a 0.6 s excerpt taken between 0.2 s and 0.8 s using a 1160 point DFT so that every 20th point matches a partial frequency. This choice leads to a clear visual presentation. The Hamming window was used in the computation.

Three partial components were extracted from the test tone. The chosen components were the first, 25th, and 55th. The order of the FIR-based ICF filter used in the analysis is 810, and the IIR-based filter consists of 30 biquads and a fractional-delay filter of order 29. In addition, the length of the pure delay is 60 samples, when the IIR filter is used. Results for the FIR filter-based ICF are shown in Fig. 8. The subfigures (a), (b), and (c) represent the extracted partials #1, #25, and #55, respectively. When compared to the results of the IIR filter-based ICF, which are shown in Fig. 9 (a), (b), and (c), it can be seen that the FIR filter performs better.

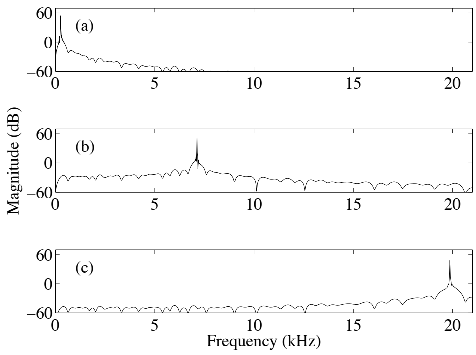

Figure 8: Results of the partial extraction in frequency domain, when an FIR-based ICF is used: (a) partial component #1, (b) partial component #25, and (c) partial component #55.

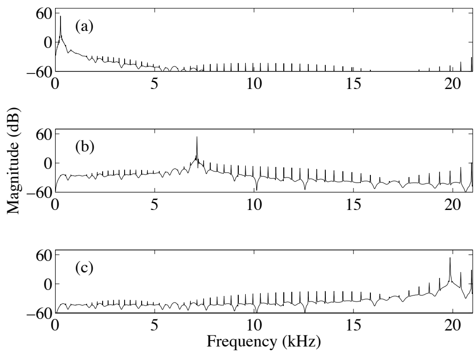

Figure 9: Results of the partial extraction in frequency domain, when an IIR-based ICF is used: (a) partial component #1, (b) partial component #25, and (c) partial component #55.

Figure 9 shows that some harmonic components are still visible in the spectra. No visible differences were found in the time domain representation, and the effect of temporal smearing is small in both cases.

4.2. Recorded Test Tones

In order to see how the proposed inharmonic ICFs perform with real tones, partials were canceled from recorded piano tones. In the following, the results for two example tones are presented: C2 (key index 16, $f_0 = 65.6206$ Hz, $B = 3.8042 \times 10^{-5}$) and D5 (key index 54, $f_0 = 587.1$ Hz, $B = 0.0012$). The result for the tone C2 is shown in Figs. 10 and 11 in the time and frequency domain, respectively. In both figures, the subfigures (a) present the original recorded tone, and the subfigures (b) and (c) refer to the residual signals that are obtained with the FIR- and IIR-based ICFs, respectively. For clarity, in Fig. 11 (b) and (c), the crosses indicate the corresponding magnitude computed exactly at the partial frequencies. The spectra are computed using the Hanning window. The order of the FIR-based ICF used in the analysis is 1200. The specification for the IIR-based ICF were as follows: the IIR filter consists of 84 biquads and a fractional-delay filter or order 29, and the length of the delay that is modeled with pure delay is 250 samples. The obtained theoretical attenuation at the partial

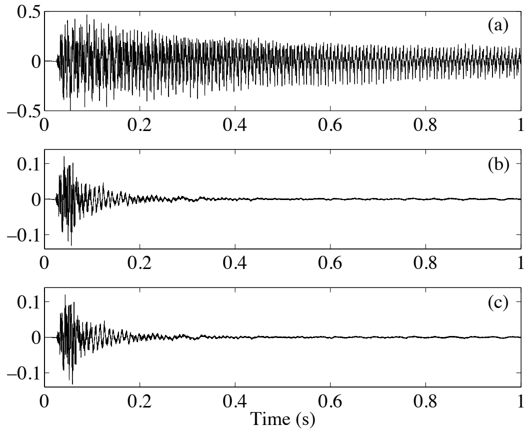

Figure 10: (a) Recorded piano tone C2 in the time domain. (b) and (c) show the residual signals that are obtained by filtering the original tone twice with the FIR- and IIR-based ICFs, respectively.

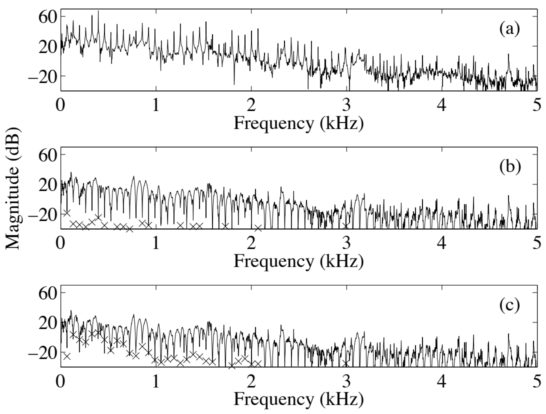

Figure 11: (a) Recorded piano tone C2 in the frequency domain. (b) and (c) show the residual signals that are obtained by filtering the original tone twice with the FIR- and IIR-based ICFs, respectively. The crosses in (b) and (c) indicate the magnitude at partial frequencies.

frequencies in the range $0 - 20$ kHz is $14 - 63$ dB with the IIR-based ICF, while with the FIR-based ICF it is possible, in theory, to obtain $98 - 122$ dB attenuation at the partial frequencies.

It was reported in [6] that in some cases filtering the signals only once is insufficient in order to get a good attenuation at the partial frequencies. This is due to the fact that the partials are not ideal peaks in the frequency domain, but they are spread around the ideal partial frequency because of beating, for example. The notches of the ICF are very narrow, and it may happen that some of the partials hit the slope of the notch rather than the bottom. Also, in this example, the signal needs to be filtered twice in the case of both ICFs. This indicates, however, that the accuracy of the phase delay specification is not very critical when real piano tones are analyzed, since the partial peaks are spread in frequency. The crosses in Fig. 11 (b) and (c) indicate, however, that the FIR-based ICF performs better, since the attenuation at the partial frequencies is better.

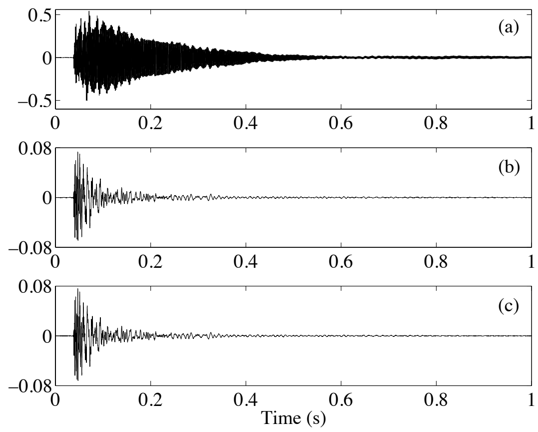

Figure 12: (a) Recorded piano tone D5 in the time domain. (b) and (c) show the residual signals that are obtained by filtering the original tone twice with the FIR- and IIR-based ICFs, respectively.

The results for the second test tone, D5, are shown in Figs. 12 and 13 in the time and frequency domains, respectively. Again, in both figures the subfigures (a) present the original recorded tone and the subfigures (b) and (c) refer to the residual signals that are obtained with the FIR- and IIR-based ICFs, respectively. The crosses in Fig. 13 (c) indicate the magnitude at partial frequencies. The corresponding crosses are not visible in Fig. 13 (b), since they fall below the visible range. The spectra are computed using the Hanning window. The order of the FIR filter is 600. The IIR filter consists of 19 biquads, and 20 samples of the delay are modeled with a delay line. In this case, however, the fractional delay filter is excluded. This is because of the error that comes from the addition of the group delay in order to achieve an area that integrates to an integer multiple of 2π. When the total area of group delay is smaller, which is the case with the high tones, the proportion of the added delay is relatively larger. In this case it is suggested that the additional delay is added without the ramp that was discussed in Sec. 2.2.1, so that the delay is evenly spread to all frequencies. To compensate the delay, the direct path of the input signal is delayed with a sufficient number of samples, in this case with two samples. With this procedure, the obtained theoretical attenuation at the partial frequencies is $14 - 32$ dB, while with the FIR-based ICF it is possible, in theory, to obtain $102 - 128$ dB attenuation.

As can be seen in Fig. 13 (b) and (c), much better attenuation is obtained when the FIR-based ICF is used, since most of the partial components are still visible in Fig. 13 (c). Some of the harmonic components can be heard after filtering the signal twice with the IIR-based ICF, while the FIR-based ICF performs well even after filtering only once. Sound examples that illustrate the effect of filtering once and twice are provided in the Internet at http://www.acoustics.hut.fi/go/DAFx08-IICF/.

4.3. Notes on Sinusoidal Modeling

Sinusoidal modeling [21, 22, 23] is one of the most powerful methods for analyzing musical tones. It is especially suitable for complex, time-varying tones. It is also a useful tool for selecting partials from musical instrument tones. On the other hand, implementing the algorithm requires quite many steps and parameter choices. First, the short-time spectrum of the signal to be analyzed

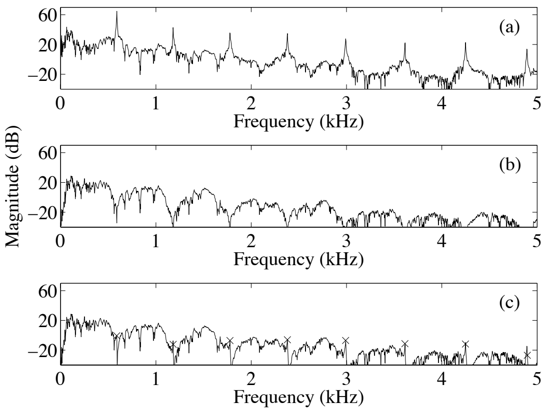

Figure 13: (a) Recorded piano tone D5 in frequency domain. (b) and (c) represent the residual signals that are obtained by filtering the original tone twice with the FIR- and IIR-based ICFs, respectively. The crosses in (c) indicate the magnitude at partial frequencies. The crosses in (b) fall below the visible range.

is computed. Then, the spectral peaks are picked based on the estimated fundamental frequency and inharmonicity coefficients. In the case of the piano, the spectral peaks may actually be clusters of partial components due to strings groups of 1-3 strings corresponding to one key, and all of these peaks need to be taken into account for good accuracy. In addition to the frequency locations, the phase and magnitude tracks need to be found. After this, the frequency, phase, and amplitude trajectories are interpolated in order to form continuous-time signals. Finally, the components are summed to form the extracted partial. If it is of interest to extract the residual signal, all the partial components need to be analyzed, synthesized, and finally be subtracted from the original tone. The parameter choices, for example in the computation of the spectrum, have an effect on the result. Especially important are the length and hop size of the short-term Fourier transform as well as the window function used. These parameters affect the accuracy in both time and frequency domains; if the attenuation of the neighboring partials is very good it might be that temporal smearing occurs. Useful software frameworks, such as CLAM [40], have been developed for using the sinusoidal modeling, which makes using the method for analyzing musical tones easier. In the proposed method, the residual signal and the partial components can be extracted with one or two filtering operations after the filter is designed.

5. CONCLUSIONS

This paper proposes two filter configurations for extracting partial components from inharmonic piano tones. The basic idea is to use an inverse comb filter to subtract a delayed version of the signal from itself. The delay must be frequency dependent in order to obtain a proper attenuation at partial frequencies. Two filters were used to produce the frequency-dependent delay: a large FIR filter that is designed using the frequency sampling method and an IIR filter proposed by Abel and Smith [17]. The two filters were compared, and it was concluded that it is possible to obtain better accuracy with the FIR-based inverse comb filter, while the computational cost of the IIR filter is much lower. In addition, the FIR filter is effortless to design and easy to control, while the nonparametric approach of the IIR filter design technique makes the design process more complicated.

The proposed techniques were tested first with a synthetic test tone. Single partials were extracted from the tone. Additionally, residual signals from two recorded piano tones were obtained using the proposed inverse comb filters. It was concluded that both of the methods are suitable as analysis tools, when the study of the features of the residual signal is of interest.

Future research includes the investigation of inverse filtering for excitation signals for digital waveguide synthesis of the piano. In order to do this, the partial decay times need to be computed and a filter matching the corresponding magnitude gain values needs to be designed. At the same time the overall phase needs to remain the same, i.e. the phase of the loop filter should be linear. The idea is tempting, since the result would provide a physical interpretation for the analysis-synthesis scheme of piano tones.

6. ACKNOWLEDGMENTS

The author would like to thank Prof. Vesa Välimäki and Mr. Stefano Zambon for their helpful comments related to this work and Dr. Balázs Bank for fruitful discussions.

7. REFERENCES

[1] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, “Physical modeling of plucked string instruments with application to real-time sound synthesis,” *J. Audio Eng. Soc.*, vol. 44, pp. 331–353, 1996.

[2] H.-M. Lehtonen, H. Penttinen, J. Rauhala, and V. Välimäki, “Analysis and modeling of piano sustain-pedal effects,” *J. Acoust. Soc. Am.*, vol. 122, no. 3, pp. 1787–1797, 2007.

[3] A. Schutz, N. Bertin, D. T. M. Slock, B. David, and R. Badeau, “Piano "forte pedal" analysis and detection,” in *Proc. 124th AES Convention*, Amsterdam, The Netherlands, May 2008.

[4] J. A. Moorer, “The optimum comb method of pitch period analysis of continuous digitized speech,” *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 22, no. 5, pp. 330–338, 1974.

[5] J. A. Moorer, “Signal processing aspects of computer music: a survey,” *Proc. IEEE*, vol. 5, no. 8, pp. 1108–1137, 1977.

[6] V. Välimäki, M. Ilmoniemi, and M. Huotilainen, “Decomposition and modification of musical instrument sounds using a fractional delay all-pass filter,” in *Proc. 6th Nordic Signal Processing Symp.*, Espoo, Finland, June 2004, pp. 208–211.

[7] V. Välimäki, H.-M. Lehtonen, and T. I. Laakso, “Musical signal analysis using the fractional-delay inverse comb filters,” in *Proc. Digital Audio Effects (DAFx-07)*, Bordeaux, France, September 2007, pp. 261–268, `http://dafx.labri.fr/main/papers/p261.pdf`.

[8] H.-M. Lehtonen, V. Välimäki, and T. I. Laakso, “Canceling and selecting partials from musical tones using fractional-delay filters,” *Computer Music J.*, vol. 32, no. 2, 2007.

[9] A. Galembo and A. Askenfelt, “Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano,” *IEEE Trans. Speech Audio Process.*, vol. 7, no. 2, pp. 197–203, March 1999.

[10] J. Rauhala, H.-M. Lehtonen, and V. Välimäki, "Fast automatic inharmonicity estimation algorithm," *J. Acous. Soc. Am. Express Letters*, vol. 121, no. 5, pp. EL184–EL189, 2007.

[11] E. Wang and B. T. G. Tan, "Application of wavelets to onset transients and inharmonicity of piano tones," *J. Audio Eng. Soc.*, vol. 56, no. 5, pp. 381–392, May 2008.

[12] S. A. Van Duyne and J. O. Smith III, "A simplified approach to modeling dispersion caused by stiffness in strings and plates," in *Proc. Int. Computer Music Conf.*, Aarhus, Denmark, September 1994, pp. 407–410.

[13] D. Rocchesso and F. Scalcon, "Accurate dispersion simulation for piano strings," in *Proc. Nordic Acoustical Meeting*, Helsinki, Finland, June 1996, pp. 407–414.

[14] I. Testa, G. Evangelista, and S. Cavaliere, "Physically inspired models for the synthesis of stiff strings with dispersive waveguides," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 7, pp. 964–977, 2004.

[15] J. Bensa, S. Bilbao, R. Kronland-Martinet, J. O. Smith III, and T. Voinier, "Computational modeling of stiff piano strings using digital waveguides and finite differences," *Acta Acust. united Acust.*, vol. 91, pp. 289–298, 2005.

[16] J. Rauhala and V. Välimäki, "Tunable dispersion filter design for piano synthesis," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 253–256, 2006.

[17] J. S. Abel and J. O. Smith III, "Robust design of very high-order allpass dispersion filters," in *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, Montreal, Quebec, Canada, Sept. 18–20, 2006, pp. 13–18, `http://www.dafx.ca/proceedings/papers/p_013.pdf`.

[18] D. Rocchesso and F. Scalcon, "Bandwidth of perceived inharmonicity for physical modeling of dispersive strings," *IEEE Trans. Speech and Audio Processing*, vol. 7, no. 5, pp. 597–601, 1999.

[19] H. Järveläinen, V. Välimäki, and M. Karjalainen, "Audibility of the timbral effects of inharmonicity in stringed musical instrument tones," *Acoustics Research Letters Online*, vol. 2, no. 3, pp. 79–84, 2001.

[20] H. Järveläinen and M. Karjalainen, "Perceptibility of inharmonicity in the acoustic guitar," *Acta Acust. united Acust.*, vol. 92, no. 5, pp. 842–847, 2006.

[21] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 774–754, 1986.

[22] X. Serra, *A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic plus Stochastic Decomposition*, Ph.D. thesis, Stanford University, California, USA, 1989, Available online at http://ccrma.stanford.edu/STANM/stanms/stanm58/.

[23] X. Serra and J. O. Smith III, "Spectral modeling synthesis: A sound analysis/synthesis based on a deterministic plus stochastic decomposition," *Computer Music J.*, vol. 14, no. 4, pp. 12–24, 1990.

[24] J. Laroche and J.-L. Meillier, "Multichannel excitation/filter modeling of percussive sounds with application to the piano," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 329–344, 1994.

[25] V. Välimäki and T. Tolonen, "Development and calibration of a guitar synthesizer," *J. Audio Eng. Soc.*, vol. 46, pp. 766–778, 1998.

[26] N. Lee, Z. Duan, and J. O. Smith III, "Excitation signal extraction for guitar tones," in *Proc. Int. Computer Music Conf.*, Copenhagen, Denmark, August 2007.

[27] G. Evangelista, "Pitch synchronous wavelet representations of speech and musical signals," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3313–3330, 1993.

[28] B. David, R. Badeau, and G. Richard, "HRHATRAC algorithm for spectral line tracking of musical signals," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Toulouse, France, May 15-19 2006, pp. 45–48.

[29] R. Badeau, B. David, and G. Richard, "High resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Processing*, vol. 54, no. 4, pp. 1341–1350, 2006.

[30] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-zooming ARMA modeling of resonant and reverberant systems," *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012–1029, 2002.

[31] P. A. A. Esquef, M. Karjalainen, and V. Välimäki, "Frequency-zooming ARMA modeling for analysis of noisy string instrument tones," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 10, pp. 953–967, 2003.

[32] J. O. Smith III, "Physical modeling using digital waveguides," *Comput. Music J.*, vol. 16, no. 4, pp. 74–91, 1992.

[33] M. Karjalainen, V. Välimäki, and T. Tolonen, "Plucked-string models: From the Karplus-Strong algorithm to digital waveguides and beyond," *Comput. Music J.*, vol. 22, no. 3, pp. 17–32, 1998.

[34] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of piano tones," *J. Acoust. Soc. Am.*, vol. 34, no. 6, pp. 749–761, 1962.

[35] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay — tools for fractional delay filter design," *IEEE Signal Process. Mag.*, vol. 13, no. 1, pp. 30–60, 1996.

[36] V. Välimäki and T. I. Laakso, "Fractional-delay filters — Design and applications," in *Nonuniform Sampling: Theory and Practice*, F. Marvasti, Ed., chapter 20, pp. 835–895. New York: Kluwer Academic/Plenum Publishers, 2001.

[37] S. C. Pei and C.-C. Tseng, "A comb filter design using fractional-sample delay," *IEEE Trans. on Circuits and Systems — Part II*, vol. 45, no. 6, pp. 649–653, 1998.

[38] T. W. Parks and C. S. Burrus, *Digital Filter Design*, John Wiley and Sons, Inc., 1987.

[39] A. Fettweis, "A simple design of maximally flat delay digital filters," *IEEE Trans. Audio and Electroacoust.*, vol. 20, no. 2, pp. 112–114, 1972.

[40] X. Amatriain, P. Arumi, and D. Garcia, "A framework for efficient and rapid development of cross-platform audio applications," *ACM Multimedia Systems J.*, vol. 14, no. 1, pp. 15–32, June 2008, http://clam.iua.upf.edu/index.html.